Consensus guidelines for validation of virtual reality surgical simulators

F. J. Carter,[1] M. P. Schijven,[2] R. Aggarwal,[3] T. Grantcharov,[4] N. K. Francis,[5] G. B. Hanna,[3] J. J. Jakimowicz[6]

[1] The Cuschieri Skills Centre, Ninewells Hospital, University of Dundee, Dundee, Scotland, DD1 9SY
[2] Department of Surgery, IJsselland Hospital, Prins Constantijnweg 2, 2906 ZC Capelle a/d Ijssel, The Netherlands
[3] Department of Surgical Oncology and Technology, Imperial College London, 10th Floor, QEQM Building, St. Mary's Hospital, Praed Street, London, W2 1NY
[4] Glostrup Hospital, Copenhagen, Denmark
[5] Ward 7, Directorate of Surgery, TUHT, Ninewells Hospital, Dundee, DD1 9SY
[6] Department of Surgery, Catharina Hospital, Michelangelolaan, 25602 ZA Eindhoven, The Netherlands

Abstract. The Work Group for Evaluation and Implementation of Simulators and Skills Training Programmes is a newly formed subgroup of the European Association of Endoscopic Surgeons (EAES). This work group undertook a review of validation evidence for surgical simulators and the resulting consensus is presented in this article. Using clinical guidelines criteria, the evidence for validation for six different simulators was rated and subsequently translated to a level of recommendation for each system. The simulators could be divided into two basic types; systems for laparoscopic general surgery and flexible gastrointestinal endoscopy. Selection of simulators for inclusion in this consensus was based on their availability and relatively widespread usage as of July 2004. Whilst level 2 recommendations were achieved for a few systems, it was clear that there was an overall lack of published validation studies with rigorous experimental methodology. Since the consensus meeting, there have been a number of new articles, system upgrades and new devices available. The work group intends to update these consensus guidelines on a regular basis, with the resulting article available on the EAES website (http://www.eaes-eur.org).

Key words: Virtual reality — Surgical training — Validation — Simulation — Consensus guidelines

The use of virtual reality (VR) surgical simulators as training tools has increased rapidly over the past few years [1–6]. There also has been a concomitant growth in the number of companies providing such systems (usually composed of a haptic interface and accompanying VR surgical simulation software) and the methods for their use worldwide. The EAES Work Group for Evaluation and Implementation of Simulators and Skills Training Programmes was set up so that a number of experts in the field could evaluate the current evidence and provide a series of guidelines. This article is the product of the first consensus meeting of the group on November 27, 2004, in London, UK, at which the topic concerning validity of VR simulators was discussed.

The place of simulation in the learning process

Before a detailed look at the concept of validation, it is useful to consider the model of learning used when training on a surgical simulator is undertaken. Currently, the most commonly used theories to explain human learning are based on constructivism. The basis of these theories is that a continuous increase in knowledge or change in behavior is brought about through learning experiences [7]. "Learning by doing" or "experiential learning" is a constructivist theory most commonly associated with Kolb [8], who described a learning cycle containing four abilities: concrete experience, reflective observation, abstract conceptualization, and active experimentation (Fig. 1).

The following illustration shows how surgical simulator training fits into this cycle. After identifying a need to learn new skills, a trainee rehearses a simulated surgical task (concrete experience), which should be followed by reflection on his or her performance (reflective
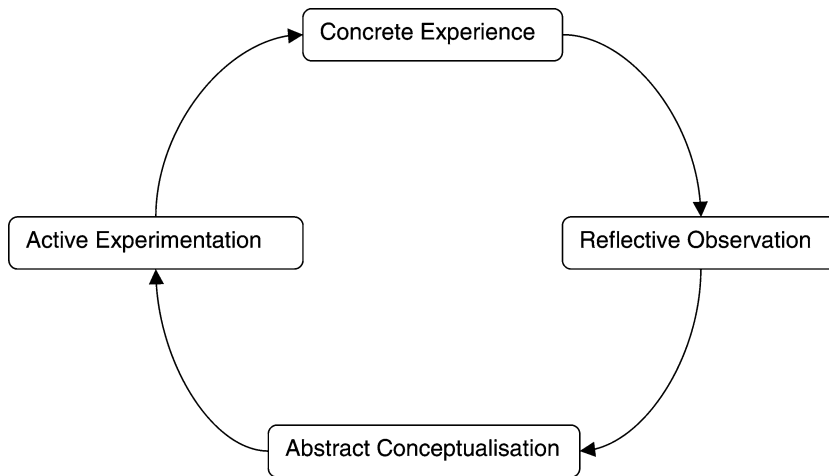
observation). Some form of task assessment and feedback to the trainee is essential to aid this reflective process. The trainee then considers ways in which his or her behavior can be modified to improve performance (abstract conceptualization) and actively experiments with these modifications by unfettered skills rehearsal on the simulator. Performance of another assessed task on the simulator moves the trainee to another set of experiences and reflections, with the cycle continuing until an acceptable level of performance is achieved.

The importance of assessment

Assessment of any task performed on a simulator, together with meaningful feedback, is a vital part of the learning process. This type of skill evaluation is a fairly informal and formative method, but it must follow the basic principles of assessment, which involve fairness, reliability, validity, and alignment with to the learning content [7].

The first important consideration is the issue of alignment of an assessment with the learning content. For example, if a learner has been instructed in a particular surgical dissection technique as part of his or her conventional training, it is important that a simulator should assess this same technique when the learner is undertaking the same procedure or task. Therefore, the validity and reliability of a learning context, such as VR surgical simulation, is of utmost importance.

Validity is defined as the extent to which an assessment instrument measures what it was designed to measure [9]. A valid VR simulator also provides an environment that closely approximates the characteristics of the environment in which the task eventually will be performed [10]. It must be able to mimic visual–spatial and real-time characteristics of the procedure, and preferably, provide realistic haptic feedback. Also, such a simulator must be able to evaluate the performance under study objectively [11].

An assessment should be able to demonstrate several forms of validity. The most basic level is that of *face validity*, in which a defined group of subjects are asked to judge the degree of resemblance between the system

under study and the real activity. *Content validity* examines the level to which the system covers the subject matter of the real activity. The degree to which the assessment can discriminate between different ability or experience levels is related to *construct* or *contrast validity*. The most powerful evidence is gained through concurrent or *predictive validity*, in which performance on the system is compared with outcomes from an established assessment method designed to measure the same skills or attributes [9, 12, 13].

The *reliability* of an evaluation instrument relates to its ability to provide consistent results with minimal errors of measurement. Test–retest reproducibility and internal consistency are the most commonly used methods for estimating internal reliability [9]. However, very few of the validation studies reviewed by the group also looked at reliability, and due to this scarcity of information, the group focused solely on validation.

The scope of this consensus document

Through literature searches and communication with simulator developers, suppliers, and other experts in the field, group members collected the available evidence on validation. This evidence then was rated according to clinical guidelines criteria [14] and thereafter translated so that a level of recommendation for each system could be established. The systems under evaluation were commercially available simulators reasonably widespread as of July 2004. The simulators for flexible endoscopy examined were Accutouch Upper and Lower GI (Immersion Medical, Gaithersburg, MD, US) and GIMentor Cyberscopy, Gastroscopy, and Colonoscopy (Simbionix, USA Corp., Cleveland OH, USA). The laparoscopic nonprocedural and hybrid simulators investigated were LapSim Basic Skills (Surgical Science, Gothenburg, Sweden), ProMIS (Haptica Ltd., Dublin, Ireland), LapMentor (Simbionix), and Procedicus MIST (Mentice, Gothenburg, Sweden). The laparoscopic procedure simulators studied were LapSim Dissection module (Surgical Science) and LapMentor LapChole module (Symbionix). Because the Xitact Corporation (Xitact SA, Morges, Switzerland) no

Table 1. Level of evidence

Level of evidence	Criteria
1a	Systematic reviews (metaanalysis) containing at least some trials of level 1b evidence, in which results of separate, independently conducted trials are consistent
1b	Randomized controlled trial of good quality and of adequate sample size (power calculation)
2a	Randomized trials of reasonable quality and/or of inadequate sample size
2b	Nonrandomized trials, comparative research (parallel cohort)
2c	Nonrandomized trial, comparative research (historical cohort, literature controls)
3	Nonrandomized, noncomparative trials, descriptive research
4	Expert opinions, including the opinion of Work Group members

Table 2. Level of recommendation based on analysis of the literature

Level of recommendation	Criteria
1	Based on one systematic review (1a) or at least two independently conducted research projects classified as 1b
2	Based on at least two independently conducted research projects classified as level 2a or 2b, within concordance
3	Based on one independently conducted research project level 2b, or at least two trials of level 3, within concordance
4	Based on one trial at level 3 or multiple expert opinions, including the opinion of Work Group members (e.g., level 4)

longer focuses on the development of simulation software, but merely emphasizes hardware development, their LapChol module was not included in this review.

Methods

The consensus development guideline was, as far as possible, based on criteria [14] and evidence resulting from literature, as previously described, according to the existing development of EAES consensus guidelines [15–17]. Relevant articles were sourced by literature search as well as communication with simulator providers and other experts in the field. Initial scoring was undertaken by individual group members, and a consensus meeting held November 2004 for discussion and agreement on levels of evidence and recommendation. Selected articles were judged according to their level of evidence according to the principle of evidence-based guideline development (Table 1).

After discussion among the group members, it was decided that abstracts, poster presentations, and personal communications could be classified only as level 4 evidence because such documents had not been scrutinized by peer review. In addition, published abstracts often lacked the detail required for a judgment on the quality of the study and could be graded only as level 4.

The outcomes from the analysis of evidence had to be interpreted to a "clinical" type of recommendation. The conclusions for each system then were categorized according to the criteria presented in Table 2 to provide transparency and remove bias.

Description of the simulators

Accutouch (Immersion Medical)

The Accutouch Lower GI Simulator consists of four modules: sigmoidoscopy, diagnostic colonoscopy, colonoscopy plus biopsy, and polypectomy. Each module contains six cases. The modules all measure end points related to procedure time, passage of the endoscope, visualization of anatomy and pathology, patient discomfort, mechanical pressures on the bowel, and usage of educational features. Increasingly complex procedures also record additional end points such as metrics related to patient sedation, tool usage, polypectomy, electrocautery, and the working channel of the endoscope.

The Accutouch Upper GI Simulator consists of two modules: one for simple diagnostic gastroscopy and one for endoscopic retrograde cholangiopancreatography (ERCP). The gastroscopy module contains six cases and measures several end points (sigmoidoscopy plus intubation metrics, record of adverse events and complications, and diagnostic instrument metrics). The ERCP module also contains six cases and measures end points similar to those of the gastroscopy module plus end points related to cannulation, pathology, fluoroscopy, and tool usage.

GIMentor (Simbionix)

The colonoscopy module contains 10 cases and measures end points relating to adverse events, procedure time, visualization of anatomy and pathology, mechanical pressures on the bowel, and use of education features. The cyberscopy module consists of generics tasks that involve aiming at simulated bubbles and catching objects in baskets. Recorded end points relate to successful actions, execution time, and economy of movement. Finally, the gastroscopy module contains 10 cases and measures end points similar to those for colonoscopy.

Procedicus MIST (Mentice)

The basic skills module consists of 12 tasks, with only the 6 basic elements included in studies to date: pick and place; pick, transfer, and place; alternate grasping; bimanual grasping and aiming; bimanual grasping plus diathermy; and a combination of the final two tasks. The end points measured are time, errors, and task efficiency.

LapSim Basic Skills (Surgical Science)

This module consists of eight tasks: camera navigation, instrument navigation, coordination, grasping, lifting and grasping, cutting, clipping, and suturing. The end points measured relate to execution time, instrument path, tissue damage, and other adverse events.

ProMIS (Haptica)

There are six ProMIS modules available, designed to train for laparoscope orientation, instrument handling, dissection, suturing and intracorporeal knot tying, diathermy, and ultrasonics. The end points measured relate to execution time, path length, and economy of movement.

LapMentor Basic Skills (Simbionix)

This module consists of five tasks: camera manipulation, hand–eye coordination, bimanual tasks, clip application, and pick and place. The

most important end points measured are execution time, number of correct hits/maneuver, accuracy rate, maintenance of horizontal view, number of camera movements, average speed of camera movements, efficiency of right and left instrument movement, path length of right and left instrument (clipper or grasper) relative to ideal path lengths, lost clips, safe clipping, and number of maneuvers.

LapMentor Lap. Chole (Symbionix)

This system consists of two full procedural tasks: clipping and division of the structures in Calot's triangle and dissection and separation of the gallbladder from the liver bed. Multiple end points are assessed, including many of the end points mentioned for the preceding module. The assessment also includes total (retraction) time, safe clipping and cutting with set distances, safe use and efficiency of cautery, accuracy rate, and percentage of completed and safe dissection.

LapSim Dissection (Surgical Science)

This system simulates the steps involved in dissection, clipping, and division of the structures in Calot's triangle and separation of the gallbladder from the liver bed. The end points recorded relate to execution time, instrument path length, tissue damage, and other adverse events.

Results

A total of 32 documents were identified as suitable for evaluation, including published articles, abstracts, posters, and personal communications. There were no metaanalyses or major randomized controlled trials addressing the issue of validation.

Level of recommendation for flexible endoscopy simulators

Concerning the level of evidence available for the flexible endoscopy simulators, Accutouch Lower GI Colonoscopy (Table 3a) has a level 2 recommendation for the diagnostic cases 1, 3, and 4 (end points: total time, percentage of mucosa seen, and scope path length) [20–24]. However there is no published evidence for the therapeutic modules. In contrast, there is scant evidence to support the validity of GIMentor Colonoscopy (Table 4a) [28–31], with contrast validation shown for unspecified cases (end points: adverse events, insertion time, and identification of pathology).

This situation is reversed for the gastroscopy simulations, with GIMentor having a level 2 recommendation for cases 1, 3, and 5 (end points: time, percentage of mucosa seen, and identification of pathology; Table 4b) [29–34]. The Accutouch Upper GI and ERCP simulations have very little published information available, with early face validation studies indicating poor validity for endoscopic appearance (Table 3b).

Level of recommendation for laparoscopic nonprocedural simulators

Considering the laparoscopic nonprocedural simulators, the strongest body of evidence is available for Procedicus MIST (Table 5), which has demonstrated contrast and concurrent validity for all six abstract tasks, resulting in a level 2 recommendation [35–42]. Although several studies have been performed on LapSim Basic Skills (Table 6), no articles have been published in peer-reviewed journals. Therefore most evidence can be given only a level 4 recommendation [43–45]. There is reasonable evidence for contrast validation for all eight tasks (unspecified end points), and some evidence of concurrent validity for instrument navigation and grasping (end point: dominant instrument path). No concurrent validation studies have been performed on Haptica ProMIS (Table 7) [46–48], but level 2c evidence exists for contrast validity in complex pick and place and sharp dissection. However, face and content validation has been demonstrated only at level 4 for clipping, cutting, and suturing. Only one validation study has been performed on Simbionix LapMentor (Table 8), showing face validity for basic skills, with the least experienced subjects rating the system the highest [49].

Level of recommendation for laparoscopic procedure simulators

Finally, the laparoscopic procedure simulations have the lowest levels of recommendation due to the lack of published validation studies. Simbionix LapMentor Laparoscopic Cholecystectomy (Table 8) has level 4 evidence for face validity, but although the choice of end points offers excellent premises for validation studies, none are currently available.

Discussion

The EAES Work group for Evaluation and Implementation of Simulators and Skills Training Programmes has undertaken the first consensus review procedure for validation of surgical simulators. A total of 32 documents concerning systems produced by five different simulation companies were examined for their level of evidence for validity. For simulation of flexible endoscopy, the highest level of recommendation is provided for Accutouch Colonoscopy and GIMentor Gastroscopy. The lowest level of recommendation is for Accutouch Upper GI modules and GIMentor Colonoscopy.

This presents a potential educator with a difficult decision. Both gastroscopy and colonoscopy must be taught and assessed with the same level of quality, and this currently would require the purchase two different simulators, according to current evidence. Clearly, there is a need for the publication of well-conducted studies based on sound experimental methods. Further work also is needed on software and hardware development for these simulators to ensure that the best quality is available throughout all modules.

Among laparoscopic nonprocedural simulators, the highest level of recommendation has been given for Procedicus MIST: level 2 for all tasks. LapSim Basic Skills has been given only a level 4 recommendation for all tasks because the studies have not yet been published

Table 3a. Level of recommendation for Accutouch Lower GI simulators

Module	Task or submodule	Type of validity				Overall level of recommendation	Comments
		Face (level)	Content (level)	Construct/Contrast (level)	Concurrent/predictive (level)		
Flexible Sigmoidoscopy	6 cases	Yes (4)	Yes (4)	Yes for total time, % mucosa seen, patient 4 (4 [18])	Yes, no specific end points (4 [19])	4	Contrast validation performed on early version of module
Colonoscopy (diagnostic)	6 cases	Yes for case 1(4 [20]) cases 3 & 4 (2b [21, 22])	Yes (4)	Yes for total time, time in red out, % mucosa seen, perforation, path length case 1 (4 [20]) cases 3 & 4 (2b [21, 22], 4 [23])	Yes, no specific end points (2a [24])	2	No published data
Colonoscopy plus biopsy	6 cases	Yes (4)	Yes (4)	—	—	—	No published data
Colonoscopy plus polypectomy	6 cases	Yes (4)	Yes (4)	—	—	—	No published data

GI, gastrointestinal

Table 3b. Level of recommendation for Accutouch Upper GI simulators

Module	Task or submodule	Type of validity				Overall level of recommendation	Comments
		Face (level)	Content (level)	Construct/Contrast (level)	Concurrent/predictive (level)		
EGD/Gastroscopy	6 cases	Yes for haptic interface (4) No for endoscopic appearance (4)	—	—	—	—	No published data
ERCP	6 cases	Yes (4) No (2c [25])	Yes (4)	—	—	—	Data is inconclusive

GI, gastrointestinal; EGD, esophagogastroduodenoscopy; ERCP, endoscopic retrograde cholangiopancreatography

Table 4a. Level of recommendation for GIMentor Colonoscopy

Module	Task or submodule	Type of validity				Overall level of recommendation	Comments
		Face (level)	Content (level)	Construct/contrast (level)	Concurrent/predictive (level)		
Colonoscopy	10 Cases	Yes for visual environment and clinical scenarios, no for haptic interface (3 [26])		Yes (4 [27]) Yes for adverse events, insertion time, and identify pathology (2b [31]) Yes for time to reach transverse colon, applied pressure and tuition required (4 [28])	—	3	Inconclusive data for validity of haptic interface

Table 4b. Level of recommendation for GIMentor Cyberscopy and Gastroscopy

Module	Task or submodule	Type of validity				Overall level of recommendation	Comments
		Face (level)	Content (level)	Construct/contrast (level)	Concurrent/predictive (level)		
Cyberscopy	Endobubble and endobasket	—	—	Yes for balloons popped (2b [29]) Yes for time, error and economy of movement (2b [30]) Yes for basket skills test (2b [31])	Correlates with MIST-VR task 6 for time, error score and economy of motion (2b [30])	2	
Gastroscopy	10 Cases	Yes for cases 1, 4, & 5 (2b [32]) Yes for unspecific cases (4 [34])		Yes for time, % mucosa & pathologies seen, cases 4 & 5 (2b [32]). Yes for adverse events, insertion time and identify pathology, case unspecified (2b [31])	Correlates with PicSOr visuospatial test for time with clear view (2b [33])	2	Trend in use of PC games relating positively to performance of GImentor

MIST-VR, Minimally Invasive Surgical Trainer in Virtual Reality; PicSOr, Pictorial Surface Orientation

Table 5. Level of recommendation for Procedicus MIST

Task or submodule	Face (level)	Content (level)	Construct/ contrast (level)	Concurrent/ predictive (level)	Overall level of recommendation	Comments
1 Pick and place	—	—	Yes (4 [35], 2b [37]) No (2b [38]) Yes for medium difficulty level (2b [41])	Yes for medium difficulty (2b [36])	2	
2 Pick, transfer, and place	—	—	Yes (4 [35]) Yes for medium difficulty level (2b [41])	Yes for medium difficulty (2b [35])	2	
3 Alternate grasping	—	—	Yes (4 [35], 2b [37]) Yes for medium difficulty level (2b [41])	Yes for medium difficulty (2b [36])	2	
4 Bimanual grasping and aiming	—	—	Yes (4 [35])No (2b [38]) Yes for medium difficulty level (2b [41])	Yes for medium difficulty (2b [36])	2	
5 Bimanual grasping plus diathermy	—	—	Yes (4 [35]) Yes for medium difficulty level (2b [41])	Yes for medium difficulty (2b [36])	2	
6 Combination of tasks 4 and 5	—	—	Yes (4 [35], 2b [37]) No (2b [38]) Yes for medium difficulty level (2b [41])	Yes for medium difficulty (2b [36])	2	
Sum of 1–6	—	—	Yes (2b [39, 40, 42])	—	3	

MIST, minimally invasive surgical trainer

Table 6. Level of recommendation for LapSim Basic Skills

Task or submodule	Face (level)	Content (level)	Construct/ contrast (level)	Concurrent/ predictive (level)	Overall level of recommendation	Comments
Camera navigation	—	—	Yes, unspecified end points (4 [45])	—	4	Abstract does not specify that this task demonstrated contrast validity
Instrument navigation	4	4	Yes for time, path length, and tissue damage (4 [44]) unspecified end points (4 [45])	Yes for dominant instrument path (4 [43])	4	As above
Coordination	—	—	Yes, unspecified end points (4 [45])	—	4	As above
Grasping	4	4	Yes, unspecified end points (4 [45])	Yes for dominant instrument path (4 [43])	4	
Lifting and grasping	4	4	Yes, unspecified end points (4 [45])	—	4	As above
Cutting	4	4	Yes for time, path length, and ripped segments (4 [44]) unspecified end points (4 [45])	—	4	
Clipping	4	4	Yes, unspecified end points (4 [45])	—	4	
Suturing	—	—	Yes, unspecified end points (4 [45])	—	4	

in peer-reviewed journals. The studies undertaken on ProMIS and LapMentor have been patchy. There is no information or evidence for some modules. This situation hopefully will change as more studies are published.

The consensus investigation also highlighted the fact that different investigators undertake validation studies in various ways. There is no uniformity of information provided to subjects, and no similar method of demonstration or familiarization of the systems in question. The questionnaires for the judgment of face validity all are designed in different ways, and authors often do not justify their method of selecting subjects into different experience groups. The most complex issue is measurement of clinical performance for predictive validation.

Table 7. Level of recommendation for ProMIS

Module	Task or submodule	Type of validity				Overall level of recommendation	Comments
		Face (level)	Content (level)	Construct/contrast (level)	Concurrent/predictive (level)		
Laparoscope orientation	Investigation (VR) Tracking (VR)	Yes, 4	—	No, 2c [46] Yes (camera navigation, 4) No (tracking, 4)	—	—	
Instrument handling	Locating and coordinating (VR) Object positioning[a] Tissue manipulation[a]	—	—	Yes, complex pick and place (2c [46]) Object positioning (4 [47])	—	4	
Dissection	Clipping[a] Cutting[a]	Yes, 4	Yes, 4 [47]	Yes, sharp dissection (2c [46], 4 [47])	—	3	
Suturing and intracorporeal knot tying	Needle–tissue interaction[a] Knot-tying[a]Ligature* Stitch placement[a]	Yes, 4	Yes, 4 [48]	—	—	4	Published study using an early version of system
Diathermy	VR task	No (4)	No (4)	—	—	—	No published information
Ultrasonics	—	—	—	—	—	—	No information provided

VR, virtual reality

Table 8. Level of recommendation for LapMentor

Module	Task or submodule	Type of validity				Overall level of recommendation	Comments
		Face (level)	Content (level)	Construct/contrast (level)	Concurrent/predictive (level)		
Basic skills	Camera manipulation, hand–eye coordination, bimanual tasks, clip applying, pick and place	Yes for unspecified tasks (4 [10])	—	—	—	—	Compared opinions of subjects of different experience, but no contrast study
Lap Chole	Clipping and cutting Calot's triangle, separating gallbladder from liver bed	Yes for unspecified tasks (4 [10])	—	—	—	—	As above

This is a topic of several research studies worldwide, and there currently exists no clear, agreed upon method for assessment of such complex skills.

It is essential to realize that not all end points measured by simulators are valid. Level 2 recommendations have been given for some modules of a few simulators, indicating that these systems are useful for formative assessment in the experiential learning cycle. For such simulators to be used for summative assessment (e.g., for selection to training programs, for certification of competence), concurrent validity must be proven for the modules and end points in question.

Some systems have been upgraded since studies have been carried out, raising the question of revalidation of such systems. The content of upgrades must be clearly explained to the end user, with good evidence provided for altering measurement of end points or type of feedback. This generally is not the case, and it is anticipated that these consensus guidelines would be an objective resource for companies considering alterations of their product.

This article presents a snapshot from an ever-expanding body of evidence in this field. The group intends to appraise new evidence regularly and update these guidelines in line with their findings. These updated documents will subsequently be available on the EAES Web site (http://www.eaes-eur.org).

References

1. Schijven M, Jakimowicz J, Broeders I, Tseng L (2005) The Eindhoven laparoscopic cholecystectomy training course: improving operating room performance using virtual reality training. Results from the first EAES-accredited virtual reality training curriculum. Surg Endosc, Accessed online 18th October 2005 at http://link.springer.de/link/service/journals/00464/index.html
2. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J (2004) Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg 91: 146–150
3. Seymour NE, Gallagher AG, Roman SA, O'Brian MK, Bansal VK (2002) Virtual reality improves operating room performance: results of a randomized, double-blind study. Ann Surg 236: 458–463
4. Hyltander A, Liljegren E, Rhodin PH, Lonroth H (2002) The transfer of basic skills learned in a laparoscopic simulator to the operating room. Surg Endosc 16: 1324–1328
5. Hamilton EC, Scott DJ, Fleming RV, Rege RV, Laycock R (2002) Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skill. Surg Endosc 16: 406–411
6. Ahlberg G, Heikkinen T, Iselius L, Leijonmarck CE, Rutqvist J (2002) Does training in a virtual reality simulator improve surgical performance? Surg Endosc 16: 126–129
7. Fry H, Ketteridge S, Marshall S (1999) Understanding student learning. In: Fry H, Ketteridge S, Marshall S (eds) A handbook for teaching and learning in higher education. Kogan Page, London
8. Kolb DA (1984) Experiential learning. Prentice-Hall, Englewood Cliffs, New Jersey
9. Murphy KR, Davidshofer CO (1998) Psychological testing: principle and applications. 4 th ed. Prentice-Hall, Upper Saddle River, New Jersey
10. Ayodeji ID, Schijven MP, Jakimowicz JJ, Bonjer HJ (2004) Face validation of the LapMentor laparoscopy trainer (Abstract) Fall meeting of the Dutch Society of Surgery, Ede, The Netherlands
11. Prystowski JB, Regehr G, Rogers DA, Loan JP, Hiemenz LL, Smith KM (1999) A virtual reality module for intravenous catheter placement. Am J Surg 177: 171–175
12. Schijven M, Jakimowicz J (2003) Construct validity: experts and novices performing on the Xitact LS500 laparoscopy simulator. Surg Endosc 17: 803–810
13. Schijven M (2005) Virtual reality simulation for laparoscopic cholecystectomy: the process of validation and implementation in the surgical curriculum outlined. Thesis, ISBN 90-9019048-1. University of Leiden, Leiden, the Netherlands
14. Andrews EJ, Redmond HP (2004) A review of clinical guidelines. Br J Surg 91: 956–964
15. Neugebauer E, Sauerland S (2000) Recommendations for evidence-based endoscopic surgery. Springer-Verlag, Paris, France
16. Fink A, Kosecoff J, Chassin M, Brook RH (1984) Consensus methods: characteristics and guidelines for use. Am J Public Health 74: 979–983
17. (NABON), N.B.O.N. and N.B.C.C.t. Netherlands (1992) Richtlijn behandeling van het mammacarcinoom (Guidelines for treatment of breast cancer). van Zuiden communications, Alphen aan de Rijn.
18. Datta VK, Mandalia M, Mackay SD, Darzi AW (2001) Evaluation and validation of a virtual reality–based flexible sigmoidoscopy trainer (Abstract). Gut 48 (Suppl 1): A97–A98
19. Garuda S, Keshavarzian A, Losurdo J, Brown MD (2002) Efficacy of a computer-assisted endoscopic simulator in training residents in flexible endoscopy (Poster). Proceedings of ACG 2002, Seattle, WA
20. Carter FJ, Steele RJC, Kennovin MF, Cuschieri A (2003) Validation of a virtual reality colonoscopy simulator using subjects of differing experience (Abstract). Proc. 1st European Endoscopic Surgical Week and 11th EAES Congress, 15th–18th June 2003, Glasgow, UK
21. Sedlack RE, Kolars JC (2003) Validation of a computer-based colonoscopy simulator. Gastrointest Endosc 57: 214–218
22. Mahmood T, Darzi A (2003) A study to validate the colonoscopy simulator: it is usefully discriminatory for more than one measurable outcome. Surg Endosc 17: 1583–1589
23. Thomas-Gibson S, Vance ME, Saunders BP (2003) Can a colonoscopy computer simulator differentiate between a novice and expert? (Abstract) Gut 52 (Suppl 1): A73
24. Sedlack RE, Kolars JC (2004) Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. Am J Gastroenterol 99: 33–37
25. Sedlack RE, Kolars JC (2003) Direct comparison of ERCP teaching models. Gastrointest Endosc 57: 886–890
26. Aalbakken L, Adamsen A, Kruse A (2000) Performance of a colonoscopy simulator: experience from a hands-on endoscopy course. Endoscopy 32: 898–900
27. Grantcharov TP, Everbusch MD, Funch-Jensen P (2004) Teaching and testing surgical skills on a VR endoscopy simulator: learning curves and impact of psychomotor training on performance in simulated colonoscopy (Poster). SAGES meeting Denver, Colorado, March–April, 2004, P179
28. Yousfi MM, Darius Sorbi MD, Baron T, Fleischer DE (2002) Flexible sigmoidoscopy: assessing endoscopic skills using a computer-based simulator (Abstract). ACG meeting, Seattle, Washington, 2002
29. Ritter EM, McClusky DA III, Lederman AB, Gallagher AG, Smith CD (2003) Objective psychomotor skills assessment of experienced and novice flexible endoscopists with a virtual reality simulator. J Gastrointest Surg 7: 871–878
30. Adamsen S, Funch-Jensen PM, Drewes AM, Rosenberg J, Grantcharov TP (2005) A comparative study of skills in virtual laparoscopy and endoscopy. Surg Endosc 19: 229–234
31. Ferlitsch A, Glauninger P, Gupper A, Schillinger M, Haefner M, Gangi A, Schoefl R (2002) Evaluation of a virtual endoscopy simulator for training in gastrointestinal endoscopy. Endoscopy 34: 698–702
32. Moorthy K, Munz Y, Jiwanji M, Bahn S, Chang A, Darzi A (2004) Validity and reliability of a virtual reality upper gastrointestinal simulator and cross validation using structured assessment of individual performance with video playback. Surg Endosc 28: 328–333
33. Enochson L, Isaksson B, Tour R, Kjellin A, Hedman L, Wredmark T, Tsai-Fellander L (2004) Visuospatial skills and computer game experience influence the performance of virtual endoscopy. J Gastrointest Surg 8: 874–880

34. Fanelli RD, Mainella MT, Justin RC, Gersin KS (2003) Initial experience using an endoscopic simulator to train residents in flexible endoscopy in a community medical center–based residency program. Flexible diagnostic and therapeutic endoscopy (Poster). Proceedings of the SAGES meeting, 12–15 March 2003, Los Angeles, California P233

35. Taffinder N, Russell RCG, McManus IC, Jansen J, Darzi A (1998) An objective assessment of surgeons' psychomotor skills: validation of the MIST-VR laparoscopic simulator. Br J Surg 85 (Supp 1): 75

36. Grantcharov TP, Rosenberg J, Pahle E, Funch-Jensen P (2001) Virtual reality computer simulation: an objective method for the evaluation of laparoscopic surgical skills. Surg Endosc 15: 242–244

37. McNatt SS, Daniel Smith C (2001) A computer-based laparoscopic skills assessment device differentiates experienced from novice laparoscopic surgeons. Surg Endosc 15: 1085–1089

38. Gallagher HJ, Allan JD, Tolley DA (2001) Spatial awareness in urologists: are they different? BJU Int 88: 666–670

39. Gallagher AG, Richie K, McClure N, McGuigan J (2001) Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. World J Surg 25: 1478–1483

40. Gallagher AG, Satava RM (2002) Virtual reality as a metric for the assessment of laparoscopic psychomotor skills: learning curves and reliability measures. Surg Endosc 16: 1746–1752

41. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J (2003) Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. Am J Surg 185: 146–149

42. Gallagher AG, Lederman AB, McGlade K, Satava RM, Smith CD (2004) Discriminative validity of the Minimally Invasive Surgical Trainer in Virtual Reality (MIST-VR) using criteria levels based on expert performance. Surg Endosc 18: 660–665

43. Carter FJ, Francis NK, Tang B, Martindale JP, Cuschieri A (2003) Validation of a virtual reality simulator with error analysis of a simulated surgical procedure (Abstract). Surg Endosc 17 (Suppl 1): S61

44. Lonroth H, Olsson JH, Johnsson E, Pazooki D (2003) Can virtual reality training tools differentiate between different levels of expertise and experience? (Abstract) Surg Endosc 17 (Suppl 1): S63

45. Duffy AJ, Hogle NJ, McCarthy H, Lew JI, Egan A, Christos PJ, Fowler DL (2003) Construct validity for the lapsim laparoscopic surgical simulator (Abstract). Surg Endosc 17 (Suppl 1): S230

46. Broe D, Ridgway PF, Johnson S, Tierney C, Conlon KC (2004) Validation of a novel hybrid surgical simulator. Proc. 12th EAES Congress 9th–12th June 2004, Barcelona, Spain

47. Hance J, Aggarwal R, Undre S, Patel H, Darzi A (2004) Evaluation of a laparoscopic video trainer with built in measures of performance. Proc. 13th SLS meeting and EndoExpo 29th September–2nd October 2004, New York

48. McClusky DAM, Van Sickle K, Gallagher AG (2004) Relationship between motion analysis, time, accuracy, and errors during performance of a laparoscopic suturing task on an augmented reality simulator'. Proc. 12th EAES Congress 9th–12th June 2004, Barcelona, Spain